BE

File copy! DO not remove

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07K — A2

(11) International Publication Number: WO 93/16099

(43) International Publication Date: 19 August 1993 (19.08.93)

(21) International Application Number: PCT/EP93/00350

(22) International Filing Date: 12 February 1993 (12.02.93)

(30) Priority data:
92102324.8 12 February 1992 (12.02.92) EP

(34) Countries for which the regional or international application was filed: DE et al.

(71) Applicant (for all designated States except US): BIOPHARM GESELLSCHAFT ZUR BIOTECHNOLOGISCHEN ENTWICKLUNG VON PHARMAKA MBH [DE/DE]; Czernyring 22, D-6900 Heidelberg (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only) : NEIDHARDT, Helge [DE/DE]; Otto-Hahn-Platz 2, D-6900 Heidelberg (DE). HÖTTEN, Gertrud [DE/DE]; Weihwiesenweg 17, D-6919 Bammental (DE).

(74) Agents: WEICKMANN, H. et al.; Kopernikusstraße 9, D-8000 München 80 (DE).

(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DNA SEQUENCES ENCODING NOVEL GROWTH/DIFFERENTIATION FACTORS

(57) Abstract

The invention provides DNA sequences encoding novel members of the TGF-β family of proteins. The TGF-β family comprises proteins which function as growth and/or differentiation factors and which are useful in medical applications. Accordingly, the invention also describes the isolation of the above-mentioned DNA sequences, the expression of the encoded proteins, the production of said proteins and pharmaceutical compositions containing said proteins.

80% homology

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

mature protein contains the most conserved sequences, especially seven cystein residues which are conserved among the family members. The TGF-ß-like proteins are multifunctional, hormonally active growth factors. They also share related biological activities such as chemotactic attraction of cells, promoting cell differentiation and their tissue-inducing capacity, such as cartilage- and bone-inducing capacity. U.S. Patent No. 5,013,649 discloses DNA sequences encoding osteo-inductive proteins termed BMP-2 proteins (bone morphogenetic protein), and U.S. patent applications serial nos. 179 101 and 179 197 disclose the BMP proteins BMP-1 and BMP-3. Furthermore, many cell types are able to synthesize TGF-ß-like proteins and virtually all cells possess TGF-ß receptors.

Taken together, these proteins show differences in their structure, leading to considerable variation in their detailed biological function. Furthermore, they are found in a wide variety of different tissues and developmental stages. Consequently, they might possess differences concerning their function in detail, for istance the required cellular physiological environment, their lifespan, their targets, their requirement for accessory factors, and their resistance to degradation. Thus, although numerous proteins exhibiting tissue-inductive, especially osteo-inductive potential are described, their natural role in the organism and, more importantly, their medical relevance must still be elucidated in detail. The occurrence of still-unknown members of the TGF-ß family relevant for osteogenesis or differentiation/induction of other tissues is strongly suspected. However, a major problem in the isolation of these new TGF-ß-like proteins is that their functions cannot yet be described precisely enough for the design of a discriminative bioassay. On the other hand, the expected nucleotide sequence homology to known members of the family would be too low to

Figure 1 shows an alignment of the amino acid sequences of MP-52 and MP-121 with some related proteins. 1a shows the alignment of MP-52 with some members of the BMP protein family starting from the first of the seven conserved cysteins; 1b shows the alignment of MP-121 with some members of the Inhibin protein family. * indicates that the amino acid is the same in all proteins compared; + indicates that the amino acid is the same in at least one of the proteins compared with MP-52 (Fig. 1a) or MP-121 (Fig. 1b).

Figure 2 shows the nucleotide sequences of the oligonucleotide primer as used in the present invention and an alignment of these sequences with known members of the TGF-ß family. M means A or C; S means C or G; R means A or G; and K means G or T. 2a depicts the sequence of the primer OD; 2b shows the sequence of the primer OID.

The present invention relates to novel TGF-ß-like proteins and provides DNA sequences contained in the corresponding genes. Such sequences include nucleotide sequences comprising the sequence

ATGAACTCCATGGACCCCGAGTCCACA and

CTTCTCAAGGCCAACACAGCTGCAGGCACC

and in particular sequences as illustrated in SEQ ID Nos. 1 and 2, allelic derivatives of said sequences and DNA sequences degenerated as a result of the genetic code for said sequences. They also include DNA sequences hybridizing under stringent conditions with the DNA sequences mentioned above and containing the following amino acid sequences:

Met-Asn-Ser-Met-Asp-Pro-Glu-Ser-Thr or

Leu-Leu-Lys-Ala-Asn-Thr-Ala-Ala-Gly-Thr.

Although said allelic, degenerate and hybridizing sequences may have structural divergencies due to naturally occurring mutations, such as small deletions or substitutions, they

9843-9847 (1990). Some typical sequence homologies, which are specific to known BMP-sequences only, were also found in the propeptide part of MP-52, whereas other parts of the precursor part of MP-52 show marked differences to BMP-precursors. The mRNA of MP-121 was detected in liver tissue, and its correspondig amino acid sequence shows homology to the amino acid sequences of the Inhibin protein chains (see Fig. 1b). cDNA sequences encoding TGF-ß-like proteins have not yet been isolated from liver tissue, probably due to a low abundance of TGF-ß specific transcripts in this tissue. In embryogenic tissue, however, sequences encoding known TGF-ß-like proteins can be found in relative abundance. The inventors have recently detected the presence of a collection of TGF-ß-like proteins in liver as well. The high background level of clones related to kown factors of this group presents the main difficulty in establishing novel TGF-ß-related sequences from these and probably other tissues. In the present invention, the cloning was carried out according to the method described below. Once the DNA sequence has been cloned, the preparation of host cells capable of producing the TGF-ß-like proteins and the production of said proteins can be easily accomplished using known recombinant DNA techniques comprising constructing the expression plasmids encoding said protein and transforming a host cell with said expression plasmid, cultivating the transformant in a suitable culture medium, and recovering the product having TGF-ß-like activity.

Thus, the invention also relates to recombinant molecules comprising DNA sequences as described above, optionally linked to an expression control sequence. Such vectors may be useful in the production of TGF-ß-like proteins in stably or transiently transformed cells. Several animal, plant, fungal and bacterial systems may be employed for the transformation and subsequent cultivation process. Preferably, expression

from bacteria such as Bacillus or Escherichia coli, from fungi such as yeast, from plants such as tobacco, potato, or Arabidopsis, and from animals, in particular vertebrate cell lines such as the Mo-, COS- or CHO cell line.

Yet another aspect of the present invention is to provide a particularly sensitive process for the isolation of DNA sequences corresponding to low abundance mRNAs in the tissues of interest. The process of the invention comprises the combination of four different steps. First, the mRNA has to be isolated and used in an amplification reaction using oligonucleotide primers. The sequence of the oligonucleotide primers contains degenerated DNA sequences derived from the amino acid sequence of proteins related to the gene of interest. This step may lead to the amplification of already known members of the gene family of interest, and these undesired sequences would therefore have to be eliminated. This object is achieved by using restriction endonucleases which are known to digest the already-analyzed members of the gene family. After treatment of the amplified DNA population with said restriction endonucleases, the remaining desired DNA sequences are isolated by gel electrophoresis and reamplified in a third step by an amplification reaction, and in a fourth step they are cloned into suitable vectors for sequencing. To increase the sensitivity and efficiency, steps two and three are repeatedly performed, at least two times in one embodiment of this process.

In a preferred embodiment, the isolation process described above is used for the isolation of DNA sequences from liver tissue. In a particularly preferred embodiment of the above-described process, one primer used for the PCR experiment is homologous to the polyA tail of the mRNA, whereas the second primer contains a gene-specific sequence. The techniques employed in carrying out the different steps of this process

might be useful for diagnostic methods.

The following examples illustrate in detail the invention disclosed, but should not be construed as limiting the invention.

Example 1
Isolation of MP-121

1.1 Total RNA was isolated from human liver tissue (40-year-old-male) by the method of Chirgwin et al., Biochemistry 18 (1979), 5294-5299. Poly A^+ RNA was separated from total RNA by oligo (dT) chromatography according to the instructions of the manufacturer (Stratagene Poly (A) Quick columns).

1.2 For the reverse transcription reaction, poly A^+ RNA (1-2.5 µg) derived from liver tissue was heated for 5 minutes to 65°C and cooled rapidly on ice. The reverse transcription reagents containing 27 U RNA guard (Pharmacia), 2.5 µg oligo $d(T)_{12-18}$ (Pharmacia) 5 x buffer (250 mM Tris/HCl pH 8.5; 50 mM $MgCl_2$; 50 mM DTT; 5 mM each dNTP; 600 mM KCl) and 20 units avian myeloblastosis virus reverse transcriptase (AMV, Boehringer Mannheim) per µg poly (A^+) RNA were added. The reaction mixture (25 µl) was incubated for 2 hours at 42°C. The liver cDNA pool was stored at -20°C.

1.3 The deoxynucleotide primers OD and OID (Fig. 2) designed to prime the amplification reaction were generated on an automated DNA-synthesizer (Biosearch). Purification was done by denaturating polyacrylamide gel electrophoresis and isolation of the main band from the gel by isotachophoresis. The oligonucleotides were designed by aligning the nucleic acid sequences of some known

each enzyme in a 2- to 12-hour reaction in a buffer recommended by the manufacturer.

1.6 Each DNA sample was fractioned by electrophoresis using a 4% agarose gel (3% FMC Nusieve agarose, Biozym and 1% agarose, BRL) in Tris borate buffer (89 mM Trisbase, 89 mM boric acid, 2 mM EDTA, pH 8). After ethidiumbromide staining uncleaved amplification products (about 200 bp; size marker was run in parallel) were excised from the gel and isolated by phenol extraction: an equal volume of phenols was added to the excised agarose, which was minced to small pieces, frozen for 10 minutes, vortexed and centrifuged. The aqueous phase was collected, the interphase reextracted by the same volume TE-buffer, centrifuged and both aqueous phases were combined. DNA was further purified twice by phenol/chloroform and once by chloroform/isoamylalcohol extraction.

1.7 After ethanol precipitation, one fourth or one fifth of the isolated DNA was reamplified using the same conditions used for the primary amplification except for diminishing the number of cycles to 13 (cycle 1: 80s 93°C/40s 52°C/40s 72°C; cycles 2-12: 60s 93°C/40s 52°C/60s 72°C; cycle 13: 60s 93°C/40s 52°C/420s 72°C). The reamplification products were purified, restricted with the same enzymes as above and the uncleaved products were isolated from agarose gels as mentioned above for the amplification products. The reamplification followed by restriction and gel isolation was repeated once.

1.8 After the last isolation from the gel, the amplification products were digested by 4 units EcoR I (Pharmacia) for 2 hours at 37°C using the buffer recommended by the manufacturer. One fourth of the restriction mixture was

likewise restricted vector pT7/T3 U19 (Pharmacia) and sequenced with the sequencing kit "Sequenase Version 2.0" (United States Biochemical Corporation). Clones were characterized by their sequence overlap to the 3' end of the known MP-121 sequence.

Example 2
Isolation of MP-52

A further cDNA sequence, MP-52, was isolated according to the above described method (Example 1) by using RNA from human embryo (8-9 weeks old) tissue. The PCR reaction contained cDNA corresponding to 20 ng of poly (A^+)RNA as starting material. The reamplification step was repeated twice for both enzyme combinations. After ligation, 24 clones from each enzyme combination were further analyzed by sequence analysis. The sample resticted by AlwN I and Sph I yielded a new sequence which was named MP-52. The other clones comprised mainly BMP6 and one BMP7 sequence. The sample restricted by Ava I, AlwN I and Tfi I contained no new sequences, but consisted mainly of BMP7 and a few Inhibin βA sequences.

The clone was completed to the 3' end according to the above described method (Example 1). The same embryo mRNA, which was used for the isolation of the first fragment of MP-52, was reverse transcribed as in Example 1. Amplification was performed using the adaptor primer (AGAATTCGCATGCCATGGTCGACG) and an internal primer (CTTGAGTACGAGGCTTTCCACTG) of the MP-52 sequence. The amplification products were reamplified using a nested adaptor primer (ATTCGCATGCCATGGTCGACGAAG) and a nested internal primer (GGAGCCCACGAATCATGCAGTCA) of the MP-52 sequence. The reamplification products were cloned after

the 3′ end of MP-52 was reverse transcribed using an internal primer of the MP-52 sequence oriented in the 5′ direction (ACAGCAGGTGGGTGGTGTGGACT). A polyA tail was appended to the 5′ end of the first strand cDNA by using terminal transferase. A two step amplification was performed first by application of a primer consisting of oligo dT and an adaptor primer (AGAATTCGCATGCCATGGTCGACGAAGC(T_{16})) and secondly an adaptor primer (AGAATTCGCATGCCATGGTCGACG) and an internal primer (CCAGCAGCCCATCCTTCTCC) of the MP-52 sequence. The amplification products were reamplified using the same adaptor primer and a nested internal primer (TCCAGGGCACTAATGTCAAACACG) of the MP-52 sequence. Consecutively the reamplification products were again reamplified using a nested adaptor primer (ATTCGCATGCCATGGTCGACGAAG) and a nested internal primer (ACTAATGTCAAACACGTACCTCTG) of the MP-52 sequence. The final reamplification products were blunt end cloned in a vector (Bluescript SK, Stratagene #212206) restricted with EcoRV. Clones were characterized by their sequence overlap to the DNA of λ 2.7.4.

Plasmid SKL 52 (H3) MP12 was deposited under number 7353 at DSM (Deutsche Sammlung von Mikroorganismen und Zellkulturen), Mascheroder Weg 1b, 3300 Braunschweig, on 10.12.1992.

Phage λ 2.7.4. was deposited under number 7387 at DSM on 13.1.1993.

SEQ ID NO: 2

SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 265 base pairs

STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULAR TYPE: cDNA to mRNA

ORIGINAL SOURCE:
ORGANISM: Human
IMMEDIATE EXPERIMENTAL SOURCE: Liver tissue

PROPERTIES: Human TGF-β-like protein (MP-121)

CATCCAGCCT GAGGGCTACG CCATGAACTT CTGCATAGGG CAGTGCCCAC TACACATAGC 60
AGGCATGCCT GGTATTGCTG CCTCCTTTCA CACTGCAGTG CTCAATCTTC TCAAGGCCAA 120
CACAGCTGCA GGCACCACTG GAGGGGGCTC ATGCTGTGTA CCCACGGCCC GGCGCCCCCT 180
GTCTCTGCTC TATTATGACA GGGACAGCAA CATTGTCAAG ACTGACATAC CTGACATGGT 240
AGTAGAGGCC TGTGGGTGCA GTTAG 265

Figure 1a

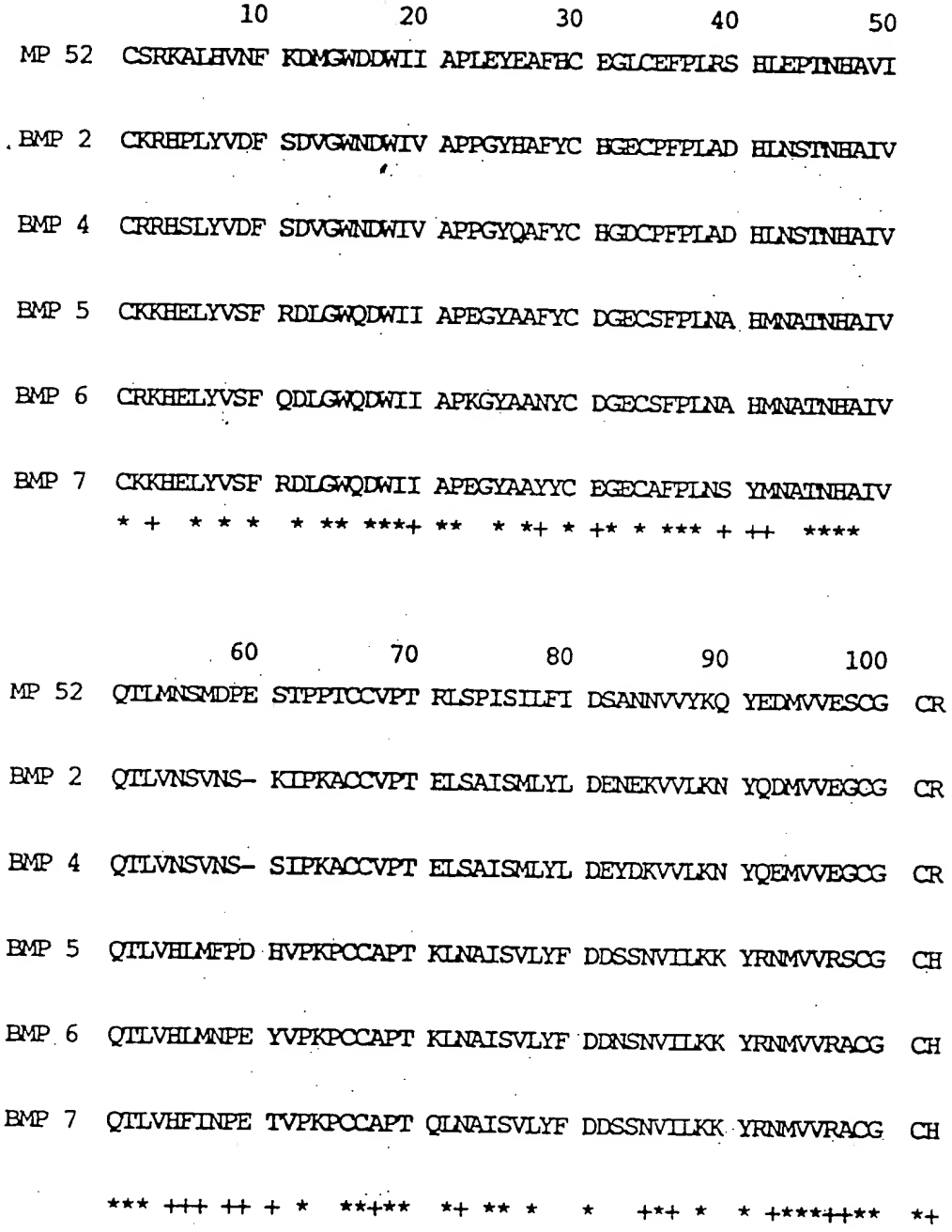
10 20 30 40 50
MP 52 CSRKALHVNF KDMGWDDWII APLEYEAFHC EGLCEFPLRS HLEPTNHAVI
BMP 2 CKRHPLYVDF SDVGWNDWIV APPGYHAFYC HGECPFPLAD HLNSTNHAIV
BMP 4 CRRHSLYVDF SDVGWNDWIV APPGYQAFYC HGDCPFPLAD HLNSTNHAIV
BMP 5 CKKHELYVSF RDLGWQDWII APEGYAAFYC DGECSFPLNA HMNATNHAIV
BMP 6 CRKHELYVSF QDLGWQDWII APKGYAANYC DGECSFPLNA HMNATNHAIV
BMP 7 CKKHELYVSF RDLGWQDWII APEGYAAYYC EGECAFPLNS YMNATNHAIV
60 70 80 90 100
MP 52 QTLMNSMDPE STPPTCCVPT RLSPISILFI DSANNVVYKQ YEDMVVESCG CR
BMP 2 QTLVNSVNS- KIPKACCVPT ELSAISMLYL DENEKVVLKN YQDMVVEGCG CR
BMP 4 QTLVNSVNS- SIPKACCVPT ELSAISMLYL DEYDKVVLKN YQEMVVEGCG CR
BMP 5 QTLVHLMFPD HVPKPCCAPT KLNAISVLYF DDSSNVILKK YRNMVVRSCG CH
BMP 6 QTLVHLMNPE YVPKPCCAPT KLNAISVLYF DDNSNVILKK YRNMVVRACG CH
BMP 7 QTLVHFINPE TVPKPCCAPT QLNAISVLYF DDSSNVILKK YRNMVVRACG CH

Figure 2a

	Eco RI Nco I
OD	ATGAATTCCCATGGACCTGGGCTGGMAKGAMTGGAT
BMP 2	ACGTGGGGTGGAATGACTGGAT
BMP 3	ATATTGGCTGGAGTGAATGGAT
BMP 4	ATGTGGGCTGGAATGACTGGAT
BMP 7	ACCTGGGCTGGCAGGACTGGAT
TGF-β1	AGGACCTCGGCTGGAAGTGGAT
TGF-β2	GGGATCTAGGGTGGAAATGGAT
TGF-β3	AGGATCTGGGCTGGAAGTGGGT
inhibin α	AGCTGGGCTGGGAACGGTGGAT
inhibin β_A	ACATCGGCTGGAATGACTGGAT
inhibin β_B	TCATCGGCTGGAACGACTGGAT

Figure 2b

	Eco RI
OID	ATGAATTCGAGCTGCGTSGGSRCACAGCA
BMP 2	GAGTTCTGTCGGGACACAGCA
BMP 3	CATCTTTTCTGGTACACAGCA
BMP 4	CAGTTCAGTGGGCACACAACA
BMP 7	GAGCTGCGTGGGCGCACAGCA
TGF-β1	CAGCGCCTGCGGCACGCAGCA
TGF-β2	TAAATCTTGGGACACGCAGCA
TGF-β3	CAGGTCCTGGGGCACGCAGCA
inhibin α	CCCTGGGAGAGCAGCACAGCA
inhibin β_A	CAGCTTGGTGGGCACACAGCA
inhibin β_B	CAGCTTGGTGGGAATGCAGCA

2. The DNA sequence according to claim 1 which is a vertebrate DNA sequence, a mammalian DNA sequence, preferably a primate, human, porcine, bovine, or rodent DNA sequence, and preferably including a rat and a mouse DNA sequence.

3. The DNA sequence according to claim 1 or 2 which is a DNA sequence comprising the nucleotides as shown in SEQ ID NO. 1.

4. The DNA sequence according to claim 1 or 2 which is a DNA sequence comprising the nucleotides as shown in SEQ ID NO. 2.

5. A recombinant DNA molecule comprising a DNA sequence according to any one of claims 1 to 4.

6. The recombinant DNA molecule according to claim 5 in which said DNA sequence is functionally linked to an expression-control sequence.

7. A host containing a recombinant DNA molecule according to claim 5 or 6.

8. The host according to claim 7 which is a bacterium, a fungus, a plant cell or an animal cell.

9. A process for the production of a protein of the TGF-ß family comprising cultivating a host according to claim 7 or 8 and recovering said TGF-ß protein from the culture.

10. A protein of the TGF-ß family encoded by a DNA sequence according to any one of claims 1 to 4.

Figure 1a

```
                    10         20         30         40         50
MP 52    CSRKALHVNF KDMGWDDWII APLEYEAFHC EGLCEFPLRS HLEPTNHAVI
BMP 2    CKRHPLYVDF SDVGWNDWIV APPGYHAFYC HGECPFPLAD HLNSTNHAIV
BMP 4    CRRHSLYVDF SDVGWNDWIV APPGYQAFYC HGDCPFPLAD HLNSTNHAIV
BMP 5    CKKHELYVSF RDLGWQDWII APEGYAAFYC DGECSFPLNA HMNATNHAIV
BMP 6    CRKHELYVSF QDLGWQDWII APKGYAANYC DGECSFPLNA HMNATNHAIV
BMP 7    CKKHELYVSF RDLGWQDWII APEGYAAYYC EGECAFPLNS YMNATNHAIV
         * +  * * *  * ** ***+ **  * *+ * +* * *** + ++  ****

                    60         70         80         90         100
MP 52    QTLMNSMDPE STPPTCCVPT RLSPISILFI DSANNVVYKQ YEDMVVESCG CR
BMP 2    QTLVNSVNS- KIPKACCVPT ELSAISMLYL DENEKVVLKN YQDMVVEGCG CR
BMP 4    QTLVNSVNS- SIPKACCVPT ELSAISMLYL DEYDKVVLKN YQEMVVEGCG CR
BMP 5    QTLVHLMFPD HVPKPCCAPT KLNAISVLYF DDSSNVILKK YRNMVVRSCG CH
BMP 6    QTLVHLMNPE YVPKPCCAPT KLNAISVLYF DDNSNVILKK YRNMVVRACG CH
BMP 7    QTLVHFINPE TVPKPCCAPT QLNAISVLYF DDSSNVILKK YRNMVVRACG CH
         *** +++ ++  + *  **+**  *+ ** *   *   +*+ *  * +*****+** *+
```

Figure 2a

	Eco RI Nco I
OD	ATGAATTCCCATGGACCTGGGCTGGMAKGAMTGGAT
BMP 2	ACGTGGGGTGGAATGACTGGAT
BMP 3	ATATTGGCTGGAGTGAATGGAT
BMP 4	ATGTGGGCTGGAATGACTGGAT
BMP 7	ACCTGGGCTGGCAGGACTGGAT
TGF-β1	AGGACCTCGGCTGGAAGTGGAT
TGF-β2	GGGATCTAGGGTGGAAATGGAT
TGF-β3	AGGATCTGGGCTGGAAGTGGGT
inhibin α	AGCTGGGCTGGGAACGGTGGAT
inhibin β_A	ACATCGGCTGGAATGACTGGAT
inhibin β_B	TCATCGGCTGGAACGACTGGAT

Figure 2b

	EcoR I
OID	ATGAATTCGAGCTGCGTSGGSRCACAGCA
BMP 2	GAGTTCTGTCGGGACACAGCA
BMP 3	CATCTTTTCTGGTACACAGCA
BMP 4	CAGTTCAGTGGGCACACAACA
BMP 7	GAGCTGCGTGGGCGCACAGCA
TGF-β1	CAGCGCCTGCGGCACGCAGCA
TGF-β2	TAAATCTTGGGACACGCAGCA
TGF-β3	CAGGTCCTGGGGCACGCAGCA
inhibin α	CCCTGGGAGAGCAGCACAGCA
inhibin β_A	CAGCTTGGTGGGCACACAGCA
inhibin β_B	CAGCTTGGTGGGAATGCAGCA